DEVIATION THEOREMS FOR PFAFFIAN SIGMOIDS

D. Yu. Grigoriev

Abstract. By a Pfaffian sigmoid with a depth d we mean a circuit with d layers in which rational operations are admitted at each layer, and to jump to the next layer one solves an ordinary differential equation of the type $v' = p(v)$ where p is a polynomial with the coefficients being the functions computed at the previous layers of the sigmoid. Thus, Pfaffian sigmoid computes Pfaffian functions (in the sense of A. Khovanskii). The deviation theorem is proved which states that for a real function $0 \not\equiv f$ computed by a Pfaffian sigmoid with a depth (or parallel complexity) d there exists an integer n such that the inequalities $(\exp(\cdots(\exp(|x|^n)\cdots)^{-1} \leq |f(x)| \leq \exp(\cdots(\exp(|x|^n)\cdots)$ hold for all $|x| \geq x_0$ for a certain x_0, where the iteration of the exponential function is taken d times. One can treat the deviation theorem as an analogue of the Liouvillean theorem (on algebraic numbers) for Pfaffian functions.

Introduction

Under Pfaffian sigmoid (cf. [MSS], [G]) with a depth d we understand a computational circuit having d layers such that at each layer the rational operations are admitted and for a jump to $(i+1)$-th layer, computing a function $w_{i+1} : \mathbb{R} \to \mathbb{R}$ at $(i+1)$-th layer is admitted, being a solution of a differential equation of the first order $w'_{i+1} = p(w_{i+1})$ (cf. [Kh]) where p is a polynomial with the coefficients computed at the previous layers of the sigmoid (see (1) and the section 1 for the exact definitions). In particular, a jump could be made by taking exp or log of a function computed at previous layers, this kind of Pfaffian sigmoids are called elementary and were considered in [G], where more generally the sigmoids were introduced in which for a jump to the next layer one can substitute a function from a previous layer into a solution of a linear ordinary differential equation with the polynomial coefficients, thus, into exp or log, in particular. Another particular case of elementary sigmoids are "standard" sigmoids (see [MSS]) where the jump is made by applying the function $(1+\exp(-x))^{-1}$. So, a function computed at $(i+1)$-th layer of a sigmoid introduced in [G], satisfies a linear ordinary differential equation with the coefficients from the previous layers $1, \dots, i$. In the present paper the corresponding differential

AMS subject classification: 68C25.

Key words and phrases: pfaffian sigmoid, deviation theorems, parallel complexity.

equation could be non-linear, but of the special form (see (1) below), providing that the computed functions are Pfaffian (see [Kh]).

The main result (see the theorem in the section 1) states that two different functions $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ computed by a Pfaffian sigmoids with a depth (or parallel complexity) d cannot be too close to each other, namely $(\exp(\cdots(\exp p(x)\cdots)))^{-1} \le |(f_1 - f_2)(x)| \le \exp(\cdots(\exp p(x)))$ for a suitable polynomial $p \in \mathbb{R}[x]$ and for all $x \ge x_0$ for some $x_0 \in \mathbb{R}$, where the number of iterations of the exponential function equals to d. This type of the results was called in [G] the deviation theorem where they were proved for the sigmoids introduced in [G] (cf. above). Also the deviation theorem could be treated as an analogue of Liouvillean theorem on the bound on the difference of two (different) algebraic numbers, for the functions computed by Pfaffian sigmoids. In particular, it gives a lower bound on the approximation of a function computed by a Pfaffian sigmoid by means of a rational function (see the corollary in the section 1). On the other hand, one could interpret the theorem and the corollary as lower bounds on the depth (so, parallel complexity) of a Pfaffian sigmoid, computing a function, provided that it has a rather good approximation by a "simple" (for example, rational) function.

The proof of the theorem is conducted by induction on the depth of a sigmoid. In the section 2 an upper bound on a function computed by a Pfaffian sigmoid is ascertained (see the lemma), in the section 3 a lower bound and thus the inductive step are proved.

1. Pfaffian sigmoids and differential fields.

Denote the field $P_0 = \mathbb{R}(X)$, then by induction on i the field P_{i+1} is generated over P_i by all the functions $w_{i+1}^{(j)} : \mathbb{R} \to \mathbb{R}$ (maybe having a finite number of singularities) satisfying first-order nonlinear differential equations of the form

$$\left(w_{i+1}^{(j)}\right)' = q\left(w_{i+1}^{(j)}\right) \tag{1}$$

where a polynomial $q(Z) \in P_i[Z]$.

According to [Kh] any function $f \in P_i$, being Pfaffian, has a finite number of singularities and roots (for $i = 1$ see also [B]). Hence for every two functions $f_1, f_2 \in P_i$, $f_1 \neq f_2$ the difference $(f_1 - f_2)(x)$ is either positive or negative everywhere on an interval $x \in [x_0, \infty)$ for a certain $x_0 \in \mathbb{R}$, we write $f_1 \succ f_2$ or $f_1 \prec f_2$ respectively. By $p_1, p_2, \dots \in \mathbb{R}[X]$ we'll denote the polynomials with the positive leading coefficients. By $\exp^{(i)} = \exp(\cdots(\exp)\cdots)$ we denote the iteration of the exponential function i times. Obviously $\exp^{(i)}(p_1)$, $(\exp^{(i)}(p_1))^{-1} \in P_i$. Now we are able to formulate the main result of the paper (cf. [G]).

THEOREM. *For any function $0 \not\equiv f \in P_i$ there exists a polynomial p_1 such that*

$$(\exp^{(i)} p_1)^{-1} \prec |f| \prec \exp^{(i)} p_1 .$$

The theorem will be proved in the next two sections, now we indicate the connections with the sigmoids (cf. [MSS]). Under a Pfaffian sigmoid with the depth d we understand a computations consisting of d layers such that a function $w_{i+1}^{(j)}$ at $(i+1)$-th layer satisfies an equation (1) where the coefficients q_ℓ, $0 \le \ell \le n$ of the polynomial $q(Z) = \sum_{0 \le \ell \le n} q_\ell Z^\ell$ are the rational functions in the functions of the form $w_i^{(j_1)}$ computed at the previous $(1, \dots, i)$ layers of the sigmoid (cf. [G]). Thus, at each layer of the Pfaffian sigmoid the rational operations are admitted and the jump from the previous layer to the next one is done by solving an equation (1). Thus, inducion on i shows that a function $w_{i+1}^{(j)}$ computed at $(i+1)$-th layer of a Pfaffian sigmoid belongs to P_{i+1}.

In [G] the elementary sigmoids were considered where the function $w_{i+1}^{(j)}$ was obtained as either $w_{i+1}^{(j)} = \exp(f_1/f_2)$ or $w_{i+1}^{(j)} = \log(f_1/f_2)$ where f_1, f_2 were the polynomials in the functions of the form $w_i^{(j_1)}$ computed at the previous $(1, \dots, i)$ layers of the elementary sigmoid. Evidently, an elementary sigmoid is a particular case of Pfaffian sigmoids. In its turn, so-called "standard" sigmoid where the jump from the previous layers to the next one is fulfilled by applying the gate function $(1 + \exp(-x))^{-1}$ ([MSS]) is a particular case of the elementary sigmoids.

An important question, how good one can approximate a function computed by a Pfaffian sigmoid by means of a rational function (cf. [G]).

COROLLARY. *If a function $w_d \in P_d$ is computed by a Pfaffian sigmoid with a depth d then for any rational function $r \in P_0$ holds $|w_d - r| \succ (\exp^{(d)}(p_2))^{-1}$ for a suitable polynomial p_2, unless $w_d = r$.*

2. Upper bound on Pfaffian function.

We start proving the theorem by induction on i. The base of induction for P_0 is obvious, let us proceed to the inductive step. In the present section we'll prove the required upper bound on the function $w_{i+1}^{(j)}$ (see (1)).

Assume by inductive hypothesis the coefficients $q_\ell \in P_i$, $q_n \neq 0$ of the polynomial q satisfy the following inequalities $(\exp^{(i)}(p_2))^{-1} \prec |q_\ell| \prec \exp^{(i)}(p_2)$, $0 \leq \ell \leq n$ for an appropriate polynomial p_2.

If $|w_{i+1}^{(j)}| \prec 4(\exp^{(i)}(p_2))^2$ is valid, the required upper bound is proved. Suppose the contrary. Then

$$\frac{1}{2}(\exp^{(i)}(p_2))^{-1} \prec \left| q_n + \frac{q_{n-1}}{w_{i+1}^{(j)}} + \cdots + \frac{q_0}{(w_{i+1}^{(j)})^n} \right| = \left| \frac{(w_{i+1}^{(j)})'}{(w_{i+1}^{(j)})^n} \right| \prec exp^{(i)}(p_2) + 1$$

If $n = 0$ then $|w_{i+1}^{(j)}(x)| \leq C_0 + \int_{x_0}^{x} (\exp^{(i)}(p_2) + 1) \prec \exp^{(i)}(p_3(x))$ for sufficiently large $x \succ x_0$ and suitable polynomial p_3 and $C_0 \in \mathbb{R}$. If $n \geq 2$ then $|(w_{i+1}^{(j)}(x))^{-n+1}| \geq (n-1) \int_{x}^{\infty} \frac{1}{2}(\exp^{(i)}(p_2))^{-1} \succ (\exp^{(i)}(p_4(x)))^{-1}$ for sufficiently large x and a suitable polynomial p_4, therefore $|w_{i+1}^{(j)}| \prec (\exp^{(i)}(p_4))$. Finally, if $n = 1$ then $\log |w_{i+1}^{(j)}(x)| \leq C_1 + \int_{x_0}^{x} (\exp^{(i)}(p_2) + 1) \prec \exp^{(i)}(p_5(x))$ for sufficiently large $x \succ x_0$ and suitable polynomial p_5 and $C_1 \in \mathbb{R}$, therefore $|w_{i+1}^{(j)}| \prec \exp^{(i+1)}(p_6)$ for a suitable polynomial p_6.

We summarize the proved upper bound in the following lemma (for $i = 0$ one can find its proof in [B]).

LEMMA. *Assume the statement of the theorem is proved for* P_i *and* $w_{i+1}^{(j)}$ *satisfies (1) where* $\deg(q) = n$. *Then*

a) *if* $n = 0$ *or* $n \geq 2$ *then* $|w_{i+1}^{(j)}| \prec \exp^{(i)}(p_7)$

b) *if* $n = 1$ *then* $|w_{i+1}^{(j)}| \prec \exp^{(i+1)}(p_7)$

for an appropriate polynomial p_7.

Remark. For any polynomial $h \in P_i[Y_1, \dots, Y_m]$ the similar upper bound as in the lemma $|h(w_{i+1}^{(j_1)}, \dots, w_{i+1}^{(j_m)})| \prec \exp^{(i+1)}(p_8)$ is valid where the functions $w_{i+1}^{(j_1)}, \dots, w_{i+1}^{(j_m)}$ satisfy similar to (1) equations, namely $(w_{i+1}^{(j_s)})' = q^{(s)}(w_{i+1}^{(j_s)})$ where $q^{(s)} \in P_i[Z]$, $1 \leq s \leq m$. This is an upper bound on a function $h(w_{i+1}^{(j_1)}, \dots, w_{i+1}^{(j_m)}) \in P_{i+1}$ required in the theorem.

3. Lower bound on Pfaffian function.

Now we proceed to proving a lower bound on $|h(w_{i+1}^{(j_1)}, \dots, w_{i+1}^{(j_m)})|$ required in the theorem. Firstly, we consider the case when $w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}$ are algebraically independent over P_i. Assume that the required lower bound is wrong, so $|h(w_{i+1}^{(j_1)}, \dots, w_{i+1}^{(j_m)})| \prec (\exp^{(i+1)}(p_\ell))^{-1}$ for all the polynomials p_ℓ. Then we say that the function $h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})$ is small. Also we assume that m is the least possible with this property. Finally, without loss of generality one can assume that the polynomial h is irreducible over P_i.

As the derivative $(h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}))' \in P_{i+1}$ is a Pfaffian function [Kh], it should be also small. One can represent $(h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}))' = \sum_{1 \leq s \leq m} \frac{\partial h}{\partial w_{i+1}^{(j_s)}} \cdot q^{(s)}(w_{i+1}^{(j_s)}) = g(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})$ for a certain polynomial $g \in P_i[Y_1, \cdots, Y_m]$.

If $h \nmid g$ in the ring $P_i[Y_1, \cdots, Y_m]$ then there exist polynomials $h_1, g_1 \in P_i[Y_1, \cdots, Y_m]$ such that $0 \not\equiv h\, h_1 + g g_1 \in P_i[Y_1, \cdots, Y_{m-1}]$ since h is irreducible. But then the function $(hh_1 + gg_1)(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_{m-1})})$ is small, applying the remark at the end of the section 2 to the polynomials h_1, g_1, that contradicts to the minimality of the choice of m.

Now suppose that $g = hg_0$ where $g_0 \in P_i[Y_1, \cdots, Y_m]$. Consider any $1 \leq s \leq m$ for which $\deg_Z(q^{(s)}) \leq 1$, then $\deg_{w_{i+1}^{(j_s)}} \left(\frac{\partial h}{\partial w_{i+1}^{(j_s)}} q^{(s)}(w_{i+1}^{(j_s)}) \right) \leq \deg_{w_{i+1}^{(j_s)}}(h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}))$

and as $\deg_{w_{i+1}^{(j_s)}} \left(\frac{\partial h}{\partial w_{i+1}^{(j_\ell)}} q^{(\ell)}(w_{i+1}^{(j_\ell)}) \right) \leq \deg_{w_{i+1}^{(j_s)}}(h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}))$ for every $\ell \neq s$, hence $w_{i+1}^{(j_s)}$ does not occur in the polynomial $g_0(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})$. If for some $1 \leq s \leq m$ $\deg_Z(q^{(s)}) \geq 2$ then lemma implies that $|w_{i+1}^{(j_s)}| \prec \exp^{(i)}(p_9)$ for an appropriate polynomial p_9. Therefore $|g_0(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})| \prec \exp^{(i)}(p_{10})$ for a certain p_{10}. Thus, $|\frac{(h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}))'}{h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})}| \prec \exp^{(i)}(p_{10})$, hence $|\log|h(w_{i+1}^{(j_1)}), \cdots, w_{i+1}^{(j_m)}| \, | \prec \exp^{(i)}(p_{11})$ and $|h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})| \succ (\exp^{(i+1)}(p_{11}))^{-1}$. This contradicts to the assumption that $h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})$ is small and proves the required in the theorem lower bound in the case when $w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}$ are algebraically independent over P_i.

In the general case choose some transcendental over P_i basis (let it be $w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_s)}$ without loss of generality) among $w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}$. Then there exists a polynomial $t(Z) = \sum_{0 \leq \ell \leq K} t^{(\ell)} Z^\ell \in P_i[w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_s)}][Z]$ where the coefficients $t^{(\ell)} \in P_i[w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_s)}]$, $0 \leq \ell \leq K$ and $t^{(0)} \not\equiv 0$, such that $t(h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})) \equiv 0$. Since we have proved that $|t^{(0)}| \succ (\exp^{(i+1)}(p_{12}))^{-1}$ and by lemma and remark after it $|t^{(\ell)}| \prec (\exp^{(i+1)}(p_{12}))$, $0 \leq \ell \leq K$ for a certain p_{12}, we obtain that $|h(w_{i+1}^{(j_1)}, \ldots, w_{i+1}^{(j_m)})| \succ \frac{1}{2}(\exp^{(i+1)}(p_{12}))^{-2}$.

This completes the proof of the inductive step in the proof of the theorem (see the beginning of the section 2) because any element in P_{i+1} can be represented as a quotient $h^{(1)}(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})/h^{(2)}(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})$ for some elements $w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)} \in P_{i+1}$ satisfying the equations of the kind (1) $(w_{i+1}^{(j_s)})' = q^{(s)}(w_{i+1}^{(j_s)}), 1 \leq s \leq m$ and for some polynomials $h^{(1)}, h^{(2)} \in P_i[Y_1, \cdots, Y_m]$. The theorem is proved.

BIBLIOGRAPHY

[B] Bellman, R., Stability theory of differential equations. McGraw-Hill Book Co., 1953..

[G] Grigoriev, D., *Deviation theorems for solutions of linear ordinary differential equations and applications to lower bounds on parallel complexity of sigmoids*, to appear.

[Kh] Khovanskii, A., *Fewnomials*, Transl. Math. Monogr. AMS, 88 (1991).

[MSS] Maass, W., Schnitger, G., Sontag, E., *On the computational power of sigmoid versus boolean threshold circuits*, Proc. 32 FOCS, IEEE, (1991), pp. 767–776.

Departments of Computer Science and Mathematics
Penn State University
State College, PA 16802 USA

On leave from
Mathematical Institute
Fontanka 27
St. Petersburg 191011 RUSSIA